NPS-55SS72071A

NAVAL POSTGRADUATE SCHOOL

Monterey, California

A SURVEY AND ANALYSIS OF HIGH DENSITY MASS STORAGE DEVICES AND SYSTEMS

by

Norman F. Schneidewind
Gordon H. Syms
Thomas L. Grainger
Robert J. Carden

July 1972

Approved for public release; distribution unlimited.

NAVAL POSTGRADUATE SCHOOL
Monterey, California

Rear Admiral Mason Freeman, USN
Superintendent

M. U. Clauser
Provost

ABSTRACT:

A survey and analysis has been made of high density mass storage systems for the Navy Fleet Material Support Office. The purpose of the project was to survey mass storage devices and systems and to select several devices for detailed analysis. Representative devices were analyzed in order to determine their suitability for various file management functions. The major conclusions of the study are the following:

1. Mass storage devices have high potential for those applications which have a requirement to store a large data base (10^{10}-10^{12} bits) on-line.
2. Mass storage devices should be considered as supplements to conventional storage devices for large data base applications, and used as part of a hierarchical storage system, rather than as replacements for conventional storage equipment.
3. Mass storage devices are not competitive with conventional storage equipment for direct access processing.
4. Erasable mass storage devices are competitive with conventional storage equipment for sequential file processing.
5. Non-erasable mass storage devices are inappropriate for high activity file processing but can be employed to advantage in archival storage applications.
6. As in the case of conventional storage units, the file activity ratio is a prime consideration in the selection of a file processing technique for mass storage. Low activity ratios favor address look-up or calculation and direct file access. High file activity ratios favor batched input and sequential file access.

This task was supported by the Fleet Material Support Office under Project Order PO-2-2099.

TABLE OF CONTENTS

I. Summary

A survey and analysis has been made of high density mass storage systems for the Navy Fleet Material Support Office. The purpose of the project was to survey mass storage devices and systems and to select several devices for detailed analysis. Representative devices were analyzed in order to determine their suitability for various file management functions. The devices and systems which were selected for analysis are the following:

. Ampex TERABIT MEMORY SYSTEM (TBM)

- Video tape mass storage, erasable media system.
- Has passed acceptance tests for first customer (undisclosed Federal agency).

. Precision Instruments UNICON 690-212

- Laser mass storage, non-erasable media system.
- First system installed at NASA Ames.

. International Video IVC-1000

- Video tape mass storage, erasable media device.
- Initial units delivered to OEM customers.

. System Development Corporation mass storage systems MMSS-1 and MMSS-2, which utilize the International Video IVC-1000 device.

- An information system designed to use IVC-1000.
- No systems have been delivered to customers.

The above were evaluated with respect to the following file management functions:

. file maintenance
. data retrieval
. report generation
. sorting

The devices were also evaluated with respect to methods of file organization and procedures for file maintenance and data retrieval.

The major conclusions of the study are the following:

1. Mass storage devices have high potential for those applications which have a requirement to store a large data base (10^{10}-10^{12} bits) on-line.

2. Mass storage devices should be considered as supplements to conventional storage devices for large data base applications, and used as part of a hierarchical storage system, rather than as replacements for conventional storage equipment.

3. Mass storage devices are not competitive with conventional storage equipment for direct access processing.

4. Erasable mass storage devices are competitive with conventional storage equipment for sequential file processing.

5. Non-erasable mass storage devices are inappropriate for high activity file processing but can be employed to advantage in archival storage applications.

6. As in the case of conventional storage units, the file activity ratio is a prime consideration in the selection of a file processing technique for mass storage. Low activity ratios favor address look-up or calculation and direct file access. High file activity ratios favor batched input and sequential file access.

II. PROJECT DESCRIPTION

A. OBJECTIVES

In recognition of the growing importance of high density storage devices and their potential for improved naval supply ADP operations, the Fleet Material Support Office (FMSO) has sponsored a Digital Recording Technology project, conducted by faculty and graduate students at the Naval Postgraduate School. The project has been conducted in two phases. The first phase, which was the subject of a previous report, involved the survey and analysis of conventional high density replacement peripheral storage devices (magnetic tapes, discs and drums). It was concluded that significant performance and cost advantages could be obtained by using devices supplied by independent manufacturers of replacement equipment. However, there was some doubt concerning the extent of user satisfaction with replacement equipment.

The second phase, which is the subject of this report, involves the survey and analysis of high density mass storage devices and systems. The devices were evaluated with respect to typical file management functions: file maintenance, data retrieval, report generation and sorting.

B. APPROACH

The approach was to search the technical literature for references to mass storage devices or systems. In all, about twelve systems or devices were identified as being in various stages of development. From this list, four devices and systems were selected for detailed analysis: Ampex TERABIT MEMORY SYSTEM (TBM), Precision Instruments UNICON 690-212, International Video IVC-1000 and the SDC MMSS-1 and MMSS-2 (uses IVC-1000 recorders).

The analysis was accomplished by reviewing technical literature and manufacturer specifications; by making a performance analysis of representative systems; and by visiting manufacturer plants to see equipment in operation and to consult with engineering and applications personnel.

Since mass storage systems are rather complex and operate differently than existing systems--often involving several stages of buffering--a large part of the effort was concentrated on analyzing the operation of the systems, using various file organizations and file maintenance procedures.

An extensive performance-cost analysis, involving numerous equations, calculations and performance-cost curves, was produced as part of this project. A future report will be issued which will provide these details.

III. INTRODUCTION TO MASS STORAGE

A. DEFINITION AND USES OF MASS STORAGE

In the literature, the most common definition of mass storage capacity is a trillion bits. A trillion bits is equivalent to 2,900 reels of standard 1600 bpi tapes, 3,500 IBM 2314 disk packs or 1,000 IBM 3300 disk packs. The possible uses of mass memory are:

1. As an extension of main memory, where long access times are acceptable.

2. For the residence of control programs and compilers. A read-only memory would be sufficient in this case.

3. For the storage of an on-line data base, possibly as part of a storage hierarchy scheme.

Manufacturers of currently available mass storage devices envision number 3 as the most likely use.

Some mass storage producers, such as Ampex, provide complete systems, including control computers, peripheral equipment and software. Other producers (International Video Corporation) are suppliers of devices to the O.E.M. market. The O.E.M. companies employ mass storage devices as part of a complete system which they sell to end users.

B. APPROACHES TO HIGH DENSITY RECORDING

Basically there are five techniques employed in achieving high density recording. They are as follows:

1. Video Magnetic Recording
   To date, video magnetic recording techniques have demonstrated packing densities of approximately 1 million bits per square inch on video tapes. Thus, for a trillion bits of data a surface area of 1 million square inches is required, implying heavy dependence on mechanical motion to control this large recording area. As a result, when video tapes are used as the recording medium, speeds of about 1 thousand inches per second are used to move the tape during searching operations.

2. Magnetic Recording
   Using magnetic tape as a recording medium, densities up to 8 thousand bits per inch per track and 16 tracks per inch have been demonstrated. This implies a somewhat larger surface area than for video recording and is also very dependent on mechanical motion. The recording surface for a trillion bits memory would be about 8 million square inches. Also, as the linear density increases, the distance from the head to tape must decrease, with possible contact recording being necessary. Higher bit and track densities necessitate smaller bit areas, thus implying smaller amounts of magnetic material to induce voltages in reading heads. Also, finer track spacings imply very difficult dimensional tolerance problems in the head positioning mechanisms.

3. <u>Optical Recording and Readout</u>
The high packing density potential of optical memories is the result of a basic principle of optics which states that an optical beam can be focused to a diffraction--limited spot whose diameter is approximately equal to the wavelength of the light. For visible light having a wavelength of approximately 0.5 microns (10^{-6} meter) the potential density of resolvable focused spots is in excess of 1 billion per square inch. In a memory, the spots must be separated by at least one or two beam diameters to uniquely define a bit, so that densities of more than 100 million bits per square inch can be expected. For this case, a trillion bit memory would only require 10,000 square inches of recording surface, which is a recording area equivalent to eight 2314 disk packs. The basic system of an optical memory will include a beam source, a beam control device, a memory medium, a beam deflector, focusing and pivoting optics, and a detector. The light source used in such a system is the laser currently used in a non-erasable mode.

4. <u>Holographic Recording</u>
When an information carrying optical beam, or signal beam, is made to intersect with a coherent optical reference beam at a pre-selected angle, which is typically between 5 and 80 degrees, then a fine-structure interference pattern results, which, when recorded forms the "hologram." A page of text could be electronically or optically composed and introduced onto the information carrying beam that propagates through an X - Y deflection system and impinges on the hologram a data storage array. The coherent optical reference beam for writing the hologram intersects the information carrying beam and thus holographically "writes" the page of text on the data storage array. To "read-out" a desired page from the holographic data storage array, an optical reference beam for reading would be directed onto the hologram and an array of photodetectors would be used to read the information from the reconstructed page of text.

   In a holographic optical memory, a page of data consisting of an array of approximately 10^4 bits is stored in a hologram having a diameter on the order of 1 millimeter. Acoustic-optic deflectors, capable of two-dimensional deflection to an array of 64 X 64 holograms with an access time approaching 1 micro-second, are presently feasible, so that approximately 4 X 10^7 bits of data are available with an access time of about 1 micro-second. About 100 square cm. of surface area is required for the above holographic array, for a recording density of 2.5 million bits per square inch. To achieve more than 10^8 bits will require mechanical motion.

5. <u>Electron Beam Recording</u>
The theoretical limit to spot size for an election beam store is of the order of 10^{-2} microns, resulting in densities of 10^{12} bits per square inch as compared to densities of 10^8 bits per square inch for optical recording.

IV. AMPEX TERABIT MEMORY SYSTEM (TBM)

A. SYSTEM DESCRIPTION

1. Memory Section

The memory section of the AMPEX TBM System is made up of Transport Modules, Transport Drivers and Data Channels.

The Transport Modules contain two tape transports with only the mechanical elements needed to move the tape. Electronic switching elements are contained in the transport but all other control logic is located in the transport driver. This design permits configurations of transport modules from one to thirty-two. Transport specifications are as follows:

TAPE SPEEDS: 5,83,248,500,1000 ips

READ WRITE DATA RATES: 6×10^6 bits/sec

TAPE REEL CAPACITY: 4.6×10^{10} bits

TRANSPORT MODULE CAPACITY: 9.2×10^{10} bits

MAXIMUM SYSTEM CAPACITY: 2.9×10^{12} bits

RECORDING MEDIA: Standard two inch magnetic video tape

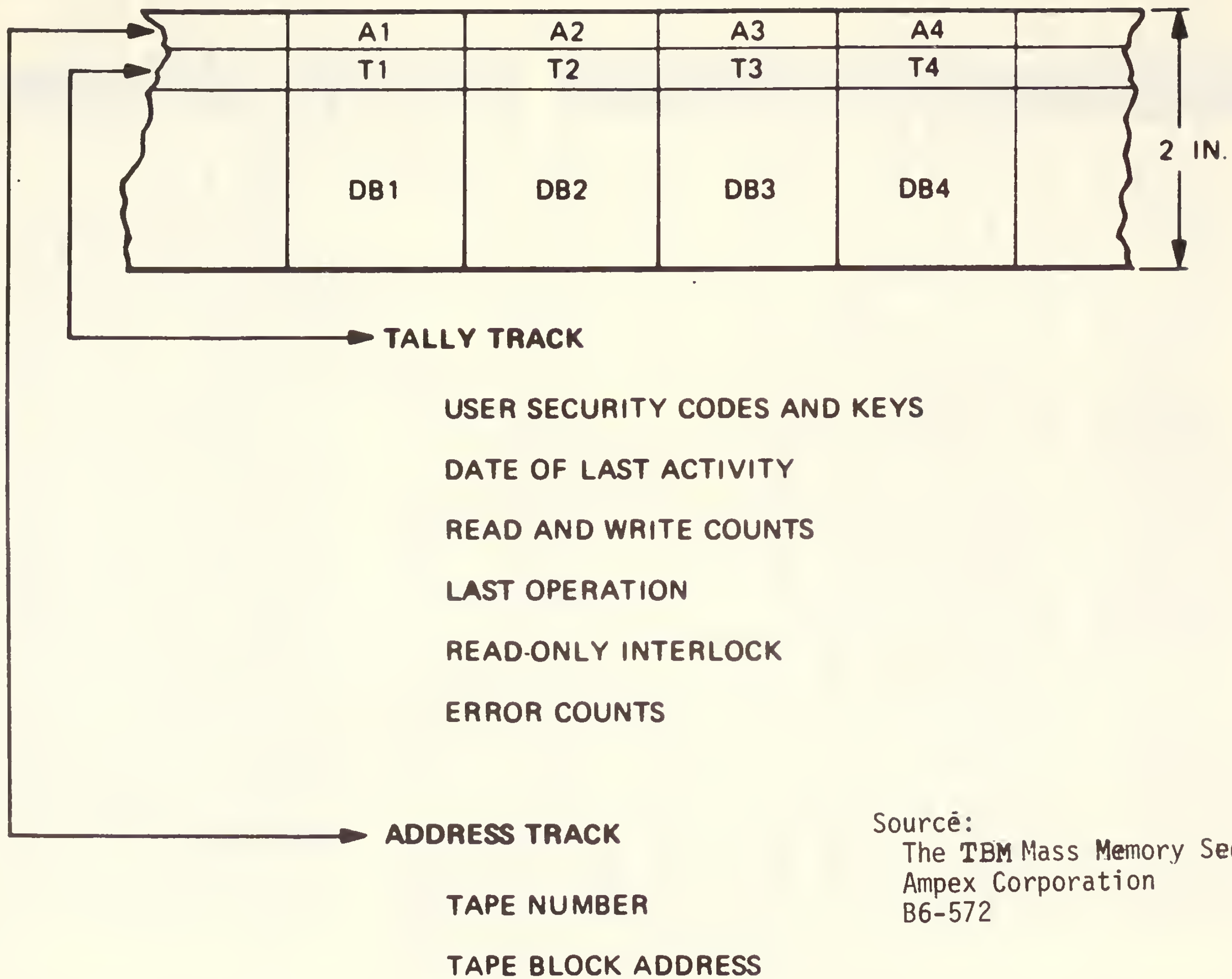
RECORDING MODE: FM

Data is recorded on the tape on transverse tracks in addressable blocks that are one inch long and hold 1 million bits. There are also two longitudinal tracks, the address track and tally track. The contents of these tracks are given in Figure IV-1. Data is read and written at 5 ips. The address track may be searched at speeds up to 1000 ips depending on the amount of tape to be searched.

The logic and electronics necessary to operate the transport are contained in the Transport Driver which is a Data General Nova mini computer. The Transport Driver controls reading, writing, and searching. The maximum number of Transport Drivers is six, permitting simultaneous read, write, and search operations that total six. The Transport Driver executes commands from the System Control Processor exclusively. It does not process user data. Switching elements allow any Transport Driver to connect to any Transport.

The Data Channel module permits simultaneous read-write operations, each at a rate of 6×10^6 bits/second. Up to six Data Channel Modules can be used permitting any combination of simultaneous reads and writes that total six. The Data Channel Module contains error detection and correction logic as well as FM encoding and decoding logic. The Data Channel also contains switching elements necessary to connect any data channel to any Transport.

TAPE FORMAT



Source:
The TBM Mass Memory Section
Ampex Corporation
B6-572

Figure IV-1 - TBM Tape Format

The Memory Section of TBM is illustrated in Figure IV-2.

2. Control Section

The control section consists of the System Control Processor (SCP), which is a PDP-11 computer, Interface Core Buffers, Staging Discs, and channel switching hardware.

The SCP controls overall functioning of the TBM system. The SCP interprets requests for search and data transfer operations initiated by the host computer. The SCP then allocates Transport Modules, Transport Drivers, Data Channels and buffers necessary to complete the operation.

Data transferred between Data Channels and Interface Core Buffers is buffered on the Staging Disc. Data is transferred in 128 K byte blocks between TBM and the Staging Disc. Data is transferred in 8 K byte segments between the Staging Disc and Interface Core Buffers. The Staging Disc provides direct access to TBM data blocks from many different files. The use of a Staging Disc also eliminates the requirement for Interface Core Buffers to be equal in size to a TBM block (128 K bytes). The Staging Disc characteristics are as follows:

CAPACITY: 9.6×10^6 bytes (75 TBM blocks)

AVERAGE ACCESS TIME: 8.5 ms

HEAD PER TRACK

CHANNELS: one or two

CHANNEL RATE: 1 million bytes/sec

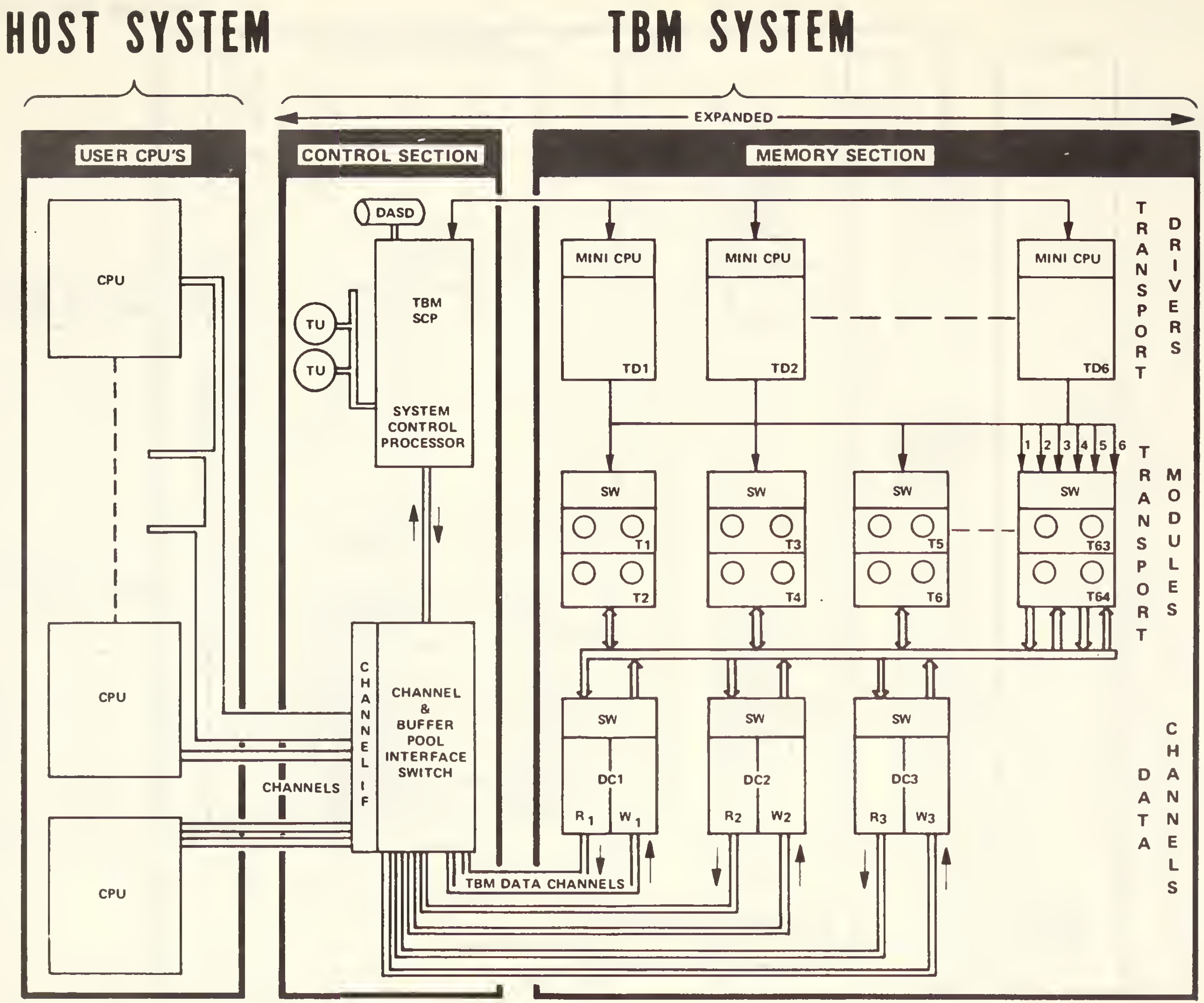
DISCS PER CONTROLLER: 1 to 4

MULTIPLE CONTROLLERS POSSIBLE

The Interface Core Buffer (IFCB) consists of a mini computer with a nominal amount of working storage for program execution plus multiple 8 K byte buffer segments up to a 128 K byte maximum which can be accessed by the host computer channel. User data is input or output by the host computer channel via the IFCB. The host computer can directly access any part of the IFCB. This feature is advantageous when only a small part of the data in a TBM block is to be processed. This feature relieves the transfer burden on host computer channels and allows for small buffers in the host computer. Switching hardware allows multipath access to Transports, Staging Discs and IFCB. The control section is illustrated in Figure IV-3.

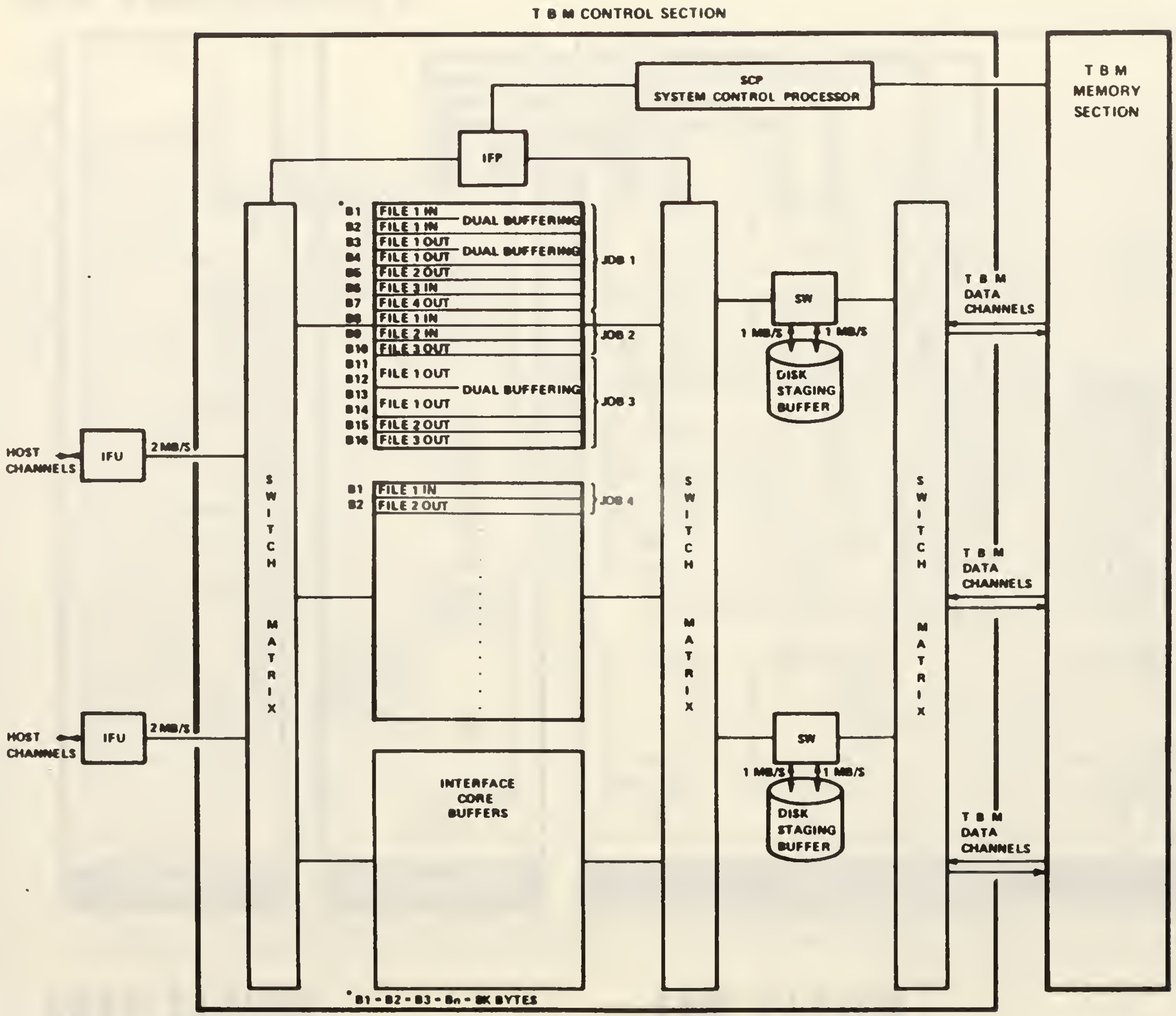
3. Off-Line Operations

System components may be placed in the off-line mode. The off-line mode is used for performing system diagnostics to locate problems, for preaddressing and testing tapes,for routine maintenance, and for file duplication for backup protection. These operations are performed with a special off-line language and do not require the host computer, the Staging Disc, or the IFCB.



Source: A Brief Description of
The Terabit Memory System
Ampex Corporation 4/7/71

Figure IV-2 - TBM Memory Section



Source: The TBM Mass
Memory System
Ampex Corporation
B6-572

Figure IV-3 - TBM Control Section

B. SOFTWARE

1. Interface Software

TBM Initiate/Terminate (TBMIT) is a software package that interfaces TBM to the IBM 360 host channel. The System Control Processor(SCP) and the Interface Processor (IFP) simulate IBM control units. TBMIT intercepts channel commands from the 360 Operating System (OS) and translates them into standard TBM commands. Job Control Language (JCL) commands must be altered to specify TBM tape numbers, block addresses, buffering required, etc. Also, OS must be set up at system generation time to process this information. TBMIT treats the application program as a subroutine and performs the housekeeping chores of locating files, opening files, and locating data prior to user program execution. Subsequent to user program execution TBMIT will close files and terminate.

At present TBMIT supports only tape sequential operations. AMPEX plans to implement direct access processing using the same approach.

To use the full capabilities of TBM, user programs must be modified to communicate TBM commands directly to the SCP. These changes will enable the user to utilize file organizations and processing methods that take advantage of TBM hardware.

2. Standard Software

Ampex indicates that it will supply an operating system which enables the SCP to control several operations. These include searching, reading, writing, control of IFCB's, control of the staging disc, and control of the switching matrix.

C. RELIABILITY AND DATA SECURITY

1. Data Error Detection and Correction

Data recording redundancy is provided by recording each bit twice, separated by approximately 3/4 inch. Data errors caused by tape defects are reduced by the separation. Two of the eight heads on the rotating head drum are always in contact with the tape. This provides for the simultaneous recording and reading of the redundant data. The data channel takes the redundant data from each of the read heads and combines the signals. Dropout in either signal causes a drop in signal level rather than a complete loss of signal. Thus no decision is required when an error occurs. The logic implemented in the channel will detect and correct any single bit error in a 955 bit block of user data. It will detect occurrances of larger error bursts. Rereads are automatically initiated if error bursts are detected.

The system was designed for an error rate of no more than 1 in 2×10^{10} bits. In acceptance tests for their first customer, AMPEX transferred 1.4×10^{12} bits of data without a hard (uncorrectable) error. In these same tests they experienced a reread rate of .37 blocks in 1000 (due to error detection and subsequent correction). Furthermore, an average 5% of the blocks on a tape required demarking because of tape defects. Experience has also indicated that a given block may be read and written about 2000 times before significant tape wear occurs. An indicator in the tally track keeps count of block accesses and sets a flag when the limit is reached. Fifty-thousand search passes (no head contact) over a given segment of tape can be expected before tape wear occurs. TBM tapes can be replaced for about $150.

2. Data Security
Protection against accidental overwriting is provided through read-only interlocks. File access protection is implemented via user codes associated with individual blocks or entire files. Backup data protection is provided by simultaneously writing data from a single IFCB to two transports. TBM tapes can also be copied independently of the host computer. The time for copying off-line is about 2 hours per TBM tape. An index file is generated for cross referencing these backup files.

3. Hardware Reliability
TBM read/write heads have an average life of 100 contact hours (range 80-120). Since there is no contact during search operations, contact hours are involved only during read/write operations. A worn-out head is refurbished at a cost of about $1500.

Component redundancy is provided via switching matricies. Any transport can be accessed by any controller and data can flow through any channel and IFCB. Thus, there is multipath access to data.

System diagnostics and maintenance programs are supplied to run in the offline mode. These programs may be run independently of the host computer. In addition, a special programming language is supplied for the off-line mode, which enables the operator to write his own test programs. The system diagnostic programs can isolate troubles down to the component level. Maintenance programs involve such functions as positioning tape loops in vacuum chambers for best tape control, pre-addressing and pre-checking new tapes, and testing the operation of servos and interval timers.

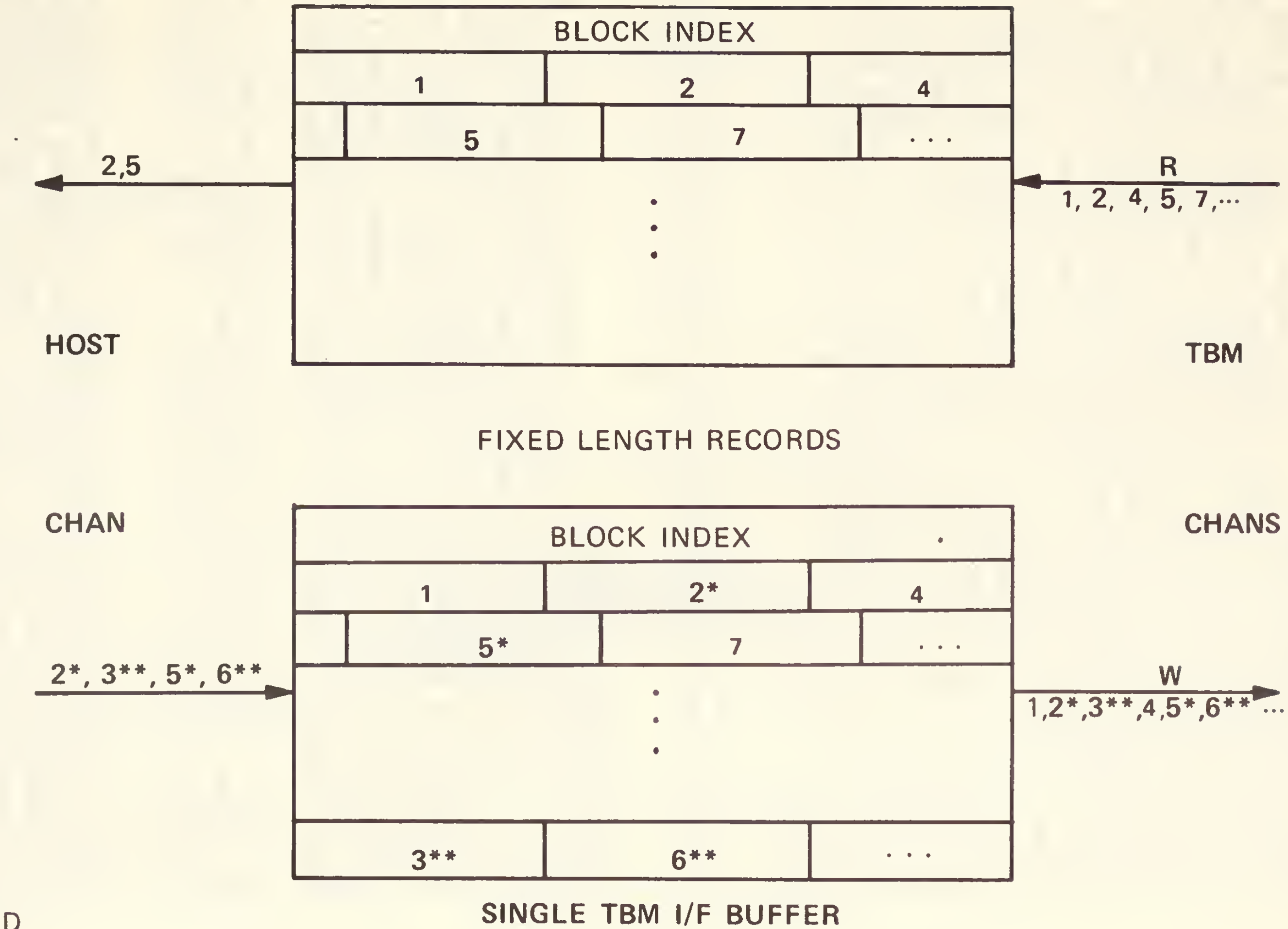
D. FILE UPDATING METHODS

1. Sequential File Updating

a. Tape-In, Tape-Out Method
This is the conventional method of updating sequential files by creating a new file in each update process. Logical records are read into IFCB's from disc where keys are examined. Records which require modification are transferred to the host computer where processing is performed. Modified records and additions to the file are transferred from the host computer back to the IFCB. Records are then transferred from the IFCB to the disc in proper sequence and written by TBM. Records to be deleted from the file are simply not transferred from the IFCB to disc.

Fixed length records are handled in the manner illustrated in Figure IV-4. Records 1,2,4,5 and 7 are read into the IFCB. Only records to be added to the file are transferred to the IFCB from the host computer. The inactive records--1, 4 and 7--remain in the IFCB. The records are written in sequence on the TBM.

ADD/UPDATE - IN PLACE



* - UPDATED
** - ADDED

Source: A Brief Description of
The Terabit Memory System
Ampex Corporation 4/7/71

Figure IV-4 - Updating Fixed Length Records
Single Buffer

Variable length records that change size are reformatted in the IFCB and are written out in sequence. This is illustrated in Figure IV-5. In this case the updated records--2 and 5--are expanded by the host computer. The added portions of the updated records and the new records--3 and 6--are transferred to the IFCB by the host computer. The records are written in sequence on the TBM.

b. <u>Tally Track Index Method</u>
Under this method, the key of the highest record in a set of TBM blocks is stored in the tally track, as shown in Figure IV-6. The Tally Track can be searched for a key by TBM at up to 83 ips. When the block set is found that contains the active record key, the block set is transferred to the disc and the active record is transferred to the IFCB. Record modifications are accomplished by the host computer and placed in the IFCB. TBM can erase the old block and rewrite the block with the record modifications. To read, erase and write a block in this manner, once the head is at the proper address, takes TBM approximately 3.5 seconds. If a record is to be deleted from the file, it is deleted from the IFCB and the block is rewritten without it. If a record is to be added to the file, it is placed in the IFCB by the host computer and transferred to TBM in its logical key sequence. Additions to the file may create overflow records from the block. In case of overflow, overflow blocks can be spaced at intervals on the same tape or on an overflow tape. An overflow index could be maintained on the host computer secondary storage.

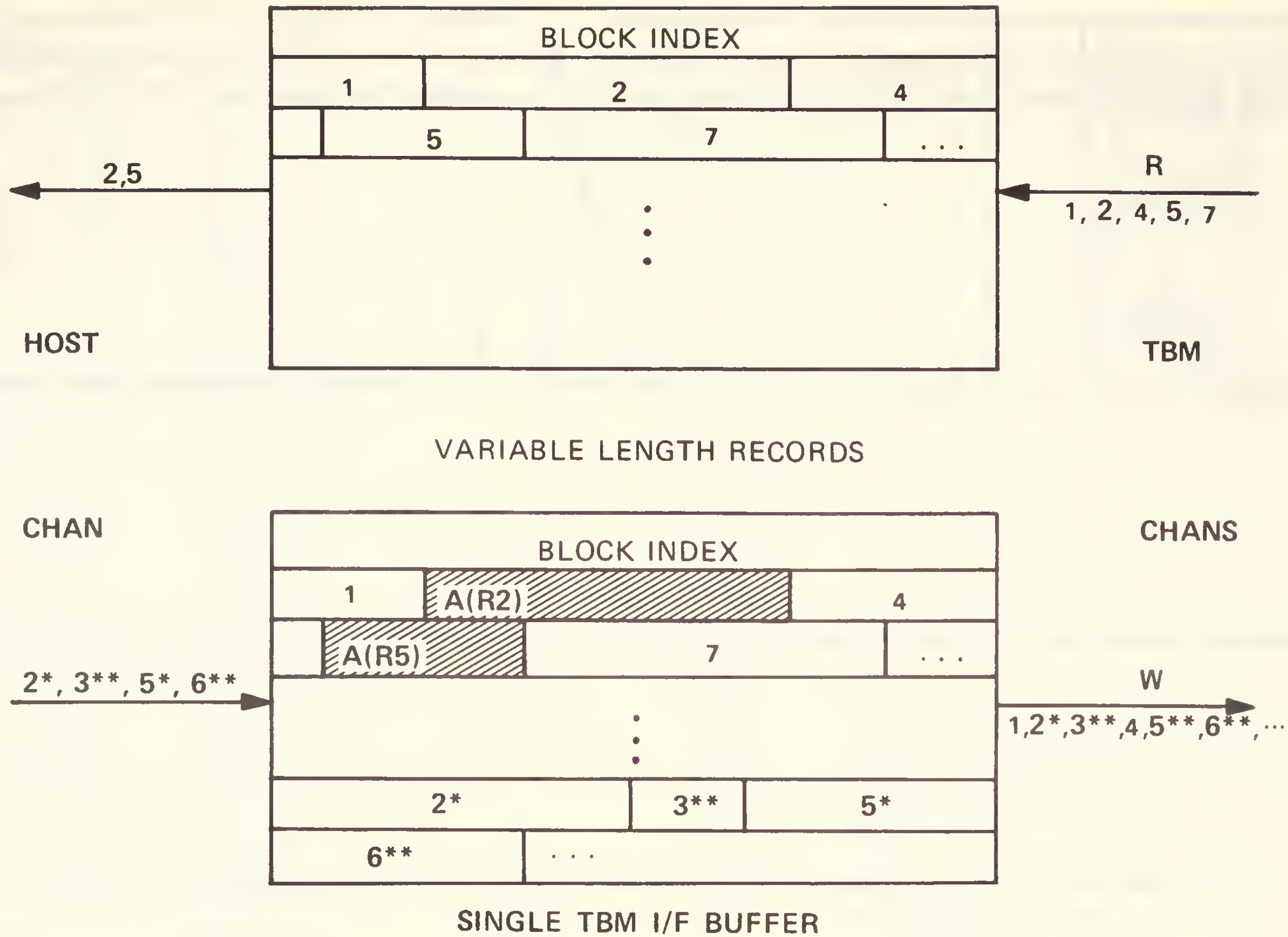
This index would list overflow keys and corresponding block addresses. This index would be searched prior to the TBM tally track search. It would also be possible to store the overflow index in the TBM block set.

c. <u>Indexed Sequential Method</u>
Using this method, data is stored sequentially on tape. Indexes can either be stored in TBM blocks or on secondary storage of the host computer. The latter method would provide the greatest speed while the former the greatest index storage capacity. These indexes would list the beginning addresses of sets of TBM blocks, and the corresponding records by key range that are found in those block sets. A primary consideration will be the size of the indexes. The massive amount of data that TBM can store will expand the indexes drastically unless only large numbers of records are included in each index entry.

For record modifications, the key of an active record is used to search the index. When a match is found, TBM tapes are searched at speeds up to 1000 ips for the indicated address. The entire block set is then transferred to disc.

Logical data records are transferred to IFCB's from disc until the active record is found. The modification is made in the host computer and returned to the IFCB. The block containing the modified record is then erased and rewritten from disc. Records to be deleted from the file are deleted from the IFCB and the TBM type block is erased and rewritten without them. Records to be added to the file are inserted in key sequence in the IFCB and transferred to disc where the block is reformatted. The TBM type block is erased and rewritten from disc. Additions may create overflow records. Overflow records may be stored in overflow blocks on the same or on

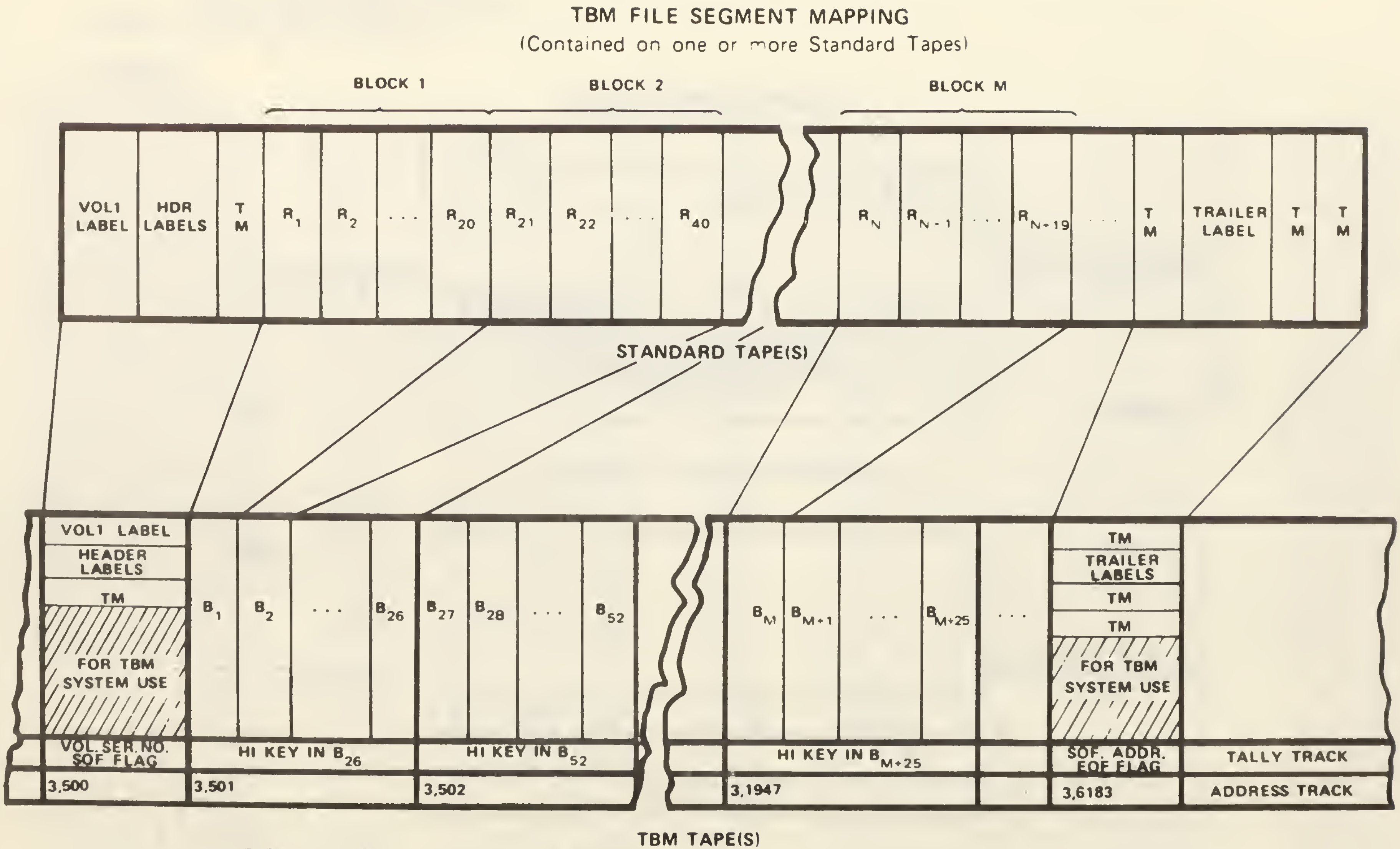
ADD/UPDATE - IN PLACE



* - UPDATED
** - ADDED

Source: A Brief Description of
The Terabit Memory System
Ampex Corporation 4/7/71

Figure IV-5 - Updating Variable Length Records
Single Buffer



Source: The TBM Mass Memory System
Ampex Corporation, B6-572

FIGURE IV-6 - Tally Track Index Method

a different TBM tape. The overflow index could be stored with the normal index.

d. Sequential With Calculated Address Method
If file records exhibit a continuous or nearly continuous set of keys, it may be possible to store records sequentially and access them through a calculated address. Large blocks of unused keys, however, will create a series of blank TBM tape blocks. The relative record address could be calculated by dividing the number of records per block into the record key. The quotient becomes the relative tape unit number and block address, while the remainder is the relative sequence of the record in the block. Record processing is performed essentially as described in the previous section.

Variable length records create problems with this method since their varying size will change the number of records per block and create overflows. One way to prevent overflows is to calculate addresses based on the minimum number of records per block (i.e. as if all records are of maximum size). This method may leave much of the TBM tape unused.

e. Sequential Across-Tapes Method
This method may be used to take advantage of the simultaneous accesses and transfers permitted on TBM. The file is in key sequence across tape drives as follows:

Tape drive	1	2	3
	Key 1	Key 2	Key 3
	Key 4	Key 5	Key 6
	Key 7	etc	

This method is advantageous when one job is to be run at a time and that job requires a large file. The simultaneity permitted by TBM can then be used to insure a continuous data flow. Updating would use the tape-in tape-out method. However, there would be several tapes in and out simultaneously with this method.

f. Efficiency
The most efficient method of updating sequential files is to batch transactions. Since the files are in sequential order, processing of presorted batches of transactions can proceed with a minimum of tape searching. The tape-in, tape-out method and the sequential-across-tapes method are the most efficient since all records will be in sequence with no overflows. The other methods which were described may require overflow blocks with overflow records physically out of sequence. With these methods, the searching and maintenance of overflow pointers or indexes will add time to the batch update process.

g. On-Demand Updating

On-demand updating involves updating transactions as they occur. The efficiency of on-demand updating depends on the speed of locating a record, given its key. With tally track index, indexed sequential and sequential with calculated address data blocks are located by an address or key and sequential searching in IFCB's is confined to a relatively few records. Address searching of TBM blocks takes place at speeds up to 1000 ips. With the tape-in, tape-out and sequential-across-tapes methods, all records must be transferred sequentially to the IFCB and searched until a key match is found. The large amount of data transfer required makes these methods relatively slow for on-demand updating. However, the search can be confined to a single physical TBM tape within a large logical file. This can be accomplished by maintaining a small index that indicates the key ranges to be found on each tape.

h. Batch Report Data Retrieval

With the tape-in, tape-out and sequential across tapes methods, data can be located only by reading records into the IFCB and matching data with selection criteria. When the selection criteria are record keys, each record must be transferred to the IFCB for a key match. All master records must be transferred in sequence until the report is complete. When the selection criteria are field data values, each record in the file must be checked. All file records must be transferred to the IFCB to be matched against the criteria. In each case selected records are transferred to the host computer for processing.

With tally track index, indexed sequential and sequential with calculated address methods, data can be located by block address when search keys are known. Only those blocks known to contain the required record need be transferred to the IFCB for a key match. Overflow record pointers may be contained in TBM blocks. To locate overflow records in this case, the record's "home" block is read into the IFCB and searched for the overflow address. Retrieval requires an additional search to the overflow block. The overflow block is transferred to the IFCB where a key match is made. The required record is then sent to the host computer for processing. Selection of records based on field data values requires a full file search. Each record must be transferred to the IFCB where its data fields are matched with selection criteria. Required records are sent to the host computer for processing. A full file search may be advantageous also when extracting records by key. This would occur for bulk reports with a high hit ratio. When the probability of finding a required record in each block is high, (most blocks are retrieved) address calculation or index searching is time consuming and unnecessary.

Considerable savings in file searching time can be made by batching reports for a single file pass. With this method records are selected for many reports as the file is passed once.

i. On-Demand Record Retrieval

On-demand record retrieval consists of satisfying randomly occuring record extraction requests as they are received. The tape-in, tape-out and sequential-across-tapes methods are not well suited to this type of retrieval. Full file searches in the IFCB's for one or a few records would be time consuming. The other three methods may be implemented to extract records by key on demand. Under these methods the address of the block or blocks containing the data can be determined. Only these blocks need be transferred to the IFCB for a key match. The required record or records are transferred to the host computer for processing.

When the selection criteria for record extraction are field data values, a full file search is necessary under all methods discussed thus far. These methods are inappropriate for on-demand retrieval. However, sequential file structures utilizing addresses (indexed sequential) can be augmented to provide a type of inverted file. Two alternatives for achieving this capability are as follows:

- A cross reference index would provide block addresses or record keys for a particular descriptor ("electrical engineers"). The index would be consulted to provide block addresses for records matching the selection criteria.

- Block addresses stored in each logical record could chain together all records having the same descriptor. A smaller index would give the entry point record key for each chain.

As with inverted files on conventional storage devices, a large amount of index maintenance is required as records are changed or moved.

Once the block address of a record satisfying selection criteria was known, the block would be accessed and transferred to the IFCB. Here each record would be matched against keys or selection criteria. Selected records would be transferred to the host computer for processing.

2. Random File Updating

With random file updating the external identifier (key) of a record is transformed into a storage address. Records are then stored at these calculated addresses. For retrieval of records for reporting or updating, the address is recalculated using the same transformation procedure. With TBM the transformation procedure could provide the address of a single block or a series of blocks (block set). The location of a specific record in the block or block set would take place in the IFCB by matching keys. The advantages of random record mapping over sequential methods are:

- Fast retrieval of individual records on demand
- Large indexes for locating records are unnecessary
- External identifiers (keys) need not conform to any sequence, or regular pattern. Keys now in use may be utilized.

Any of the randomizing hashing techniques used for random files on conventional storage appear to be feasible for TBM. Some of these include division by a prime, digit analysis, folding and radix transformation. The choice of a method will depend on results obtained in terms of packing factors, and overflows.

a. <u>On-Demand Updating</u>

On-demand updating consists of processing updates as they are received by the system. Given the key of the active record, the address is calculated by whatever hashing scheme is used in the host computer. TBM then accesses the address and the block or block set is transferred to disc. The individual records are then transferred to the IFCB where a key match takes place. When the active record is found, it is sent to the host computer for processing. If the record is to be modified, it is changed in the host computer and transferred to the IFCB. The block is then reformatted on the disc to include the change. The TBM tape block is erased and rewritten from disc. If the record is to be dropped from the file, it is deleted from the IFCB and the TBM block is erased and rewritten without it. If a record is to be added to the file, it is inserted in the IFCB. The TBM tape block is erased and rewritten from disc to include the new record.

Additions and modifications to records in a file may create an overflow from the calculated address. This occurs because it is highly probable for many records to calculate to the same address. Overflows will increase as the packing factor (storage utilization) increases. Methods of handling overflows in random files on conventional devices appear to be appropriate for TBM.

1) <u>Progressive Overflow</u>

With this method, an overflow record is stored in the next sequentially available block that has room. Thus, when an overflow occurred, the next block would be transferred to disc. The overflow record would be included in that block. The TBM tape block would be erased and rewritten including the overflow record. The tally track could keep a "full-block" indicator to minimize accesses to blocks that could not accept the overflow record. The search procedure for records that used this technique would begin at the calculated block address (home block). If the record were not found, search would continue sequentially through following blocks until a key match was found.

2) <u>Chaining Method</u>

Overflow records could also be handled by putting the address of an overflow block in each block or block set. These overflow blocks could be at intervals on the same TBM tape or could be on different tapes. When overflow occurred, the overflow block whose address appeared in the "home-block" would be accessed The overflow record would then be written in the overflow block. Search procedure would access the home block first. If the record were not found, the overflow block would then be searched by using the address contained in the home block.

3) File Organization by Activity
Activity loading is used to minimize searching time. Records are loaded on the file in activity order, high activity first. TBM has a command which returns a tape file to its midpoint, after an access. This cuts the expected search distance to 1/4 the tape, as compared to 1/3 the tape for random positioning and 1/2 the tape for end point positioning. To capitalize on this feature, a scheme which loaded high activity records near the midpoint would further cut search distance and time.

b. Batch Updating
Significant savings in file maintenance time can be gained by batching updates for a random file organization. The addresses of the active records are determined by the appropriate hash technique in the host computer. Addresses are then sorted sequentially by TBM tape reel. TBM blocks on a given tape are then searched sequentially by the sorted addresses. Search time and distance per update decreases as the transaction batch size increases. The logic for processing additions, deletions, and modifications is the same as the on-demand case. Overflows can be handled by the same methods discussed for the on-demand update case.

c. On-Demand Record Retrieval
Retrieval of records on-demand, given the record key, uses the same logic as retrieval of records for on-demand updating. TBM block addresses are determined. The block is then transferred, record at a time, to the IFCB where a key match is made. Matching records are transferred to the host computer for processing.

Retrieval of records on-demand based on data field values in the record, requires transfer of all file records to the IFCB. Each record is matched against the data field selection criteria in the IFCB. Selected records are transferred to the host computer for processing. This method is not recommended for on-demand processing. Faster retrieval of records on-demand, based on data field values, can be provided with inverted files.

d. Batch Report Data Retrieval
Preparation of batch reports, given the record keys, is similar to batch updating with random files. Report record addresses are determined from record keys by the appropriate hashing technique. The addresses are sorted sequentially within each TBM tape file. Active blocks are read into the IFCB a record at a time. In the IFCB, a key match selects out desired records for transfer to the host computer. Several reports could be batched together under this method. Each search of a TBM tape could then retrieve data for several reports. This technique could significantly reduce tape passing time.

Batch reports with field data values as selection criteria can be prepared by searching the entire file. Each record must be matched to the selection criteria in the IFCB. By batching several reports of this type records could be selected for several reports with one file pass. Batch reports with field data values as selection criteria could also be prepared using inverted file methods.

V. UNICON MASS MEMORY SYSTEM

A. SYSTEM DESCRIPTION

1. Hardware Components

Precision Instruments offer a Laser Mass Memory System, the UNICON 690-212, which stores large static or semi-static data records on a permanent medium. The UNICON 690-212 is a device which uses a precisely focused laser beam to vaporize (burn) minute holes (approximately 4 by 3 micrometers) in the metallic surface of the data strip. The laser is modulated so that it is turned on to burn a hole for writing a "1" and turned off for a "0". In vaporizing bits, burn time is approximately 100 nanoseconds within a bit-cell time of 200 nanoseconds. During the writing process, light is reflected from the data strip, and the return beam is monitored in real time to provide a read-while-write data verification capability. When reading data, the incident laser light is reduced in power to avoid burning holes in the metallic surface. The reflected light is monitored to read whether a 1 or 0 is recorded.

The recording medium of the UNICON is a 31.25 inch by 4.75 inch strip, consisting of a thin metallic coating on a polyester base. Of this strip, 31 inches X 3.5 inches is used for data recording. There are approximately 11,440 tracks recorded logitudinally down the strip for a strip capacity of 1.6 billion bits of data. There are 25 strips per pack and 18 packs per carousel for a total system on-line capacity of 0.7×10^{12} bits. There is one carousel per system which is horizontally rotated between two read/write units.

The UNICON consists of two basic units: a Laser Recorder Unit which performs the write/read functions and the physical retrieval of data strips from the on-line file, and a Recorder Control Unit which interfaces with the host computer and serves as the memory system controller.

The Laser Recorder Unit is composed of two read/record units, each of which has an independent simultaneous read/write capability. It is between these two units that the carousel is rotated in order to load any selected strip for a read/write operation. The elements of the read/record unit are:

(1) a 10 inch diameter rotatable drum which operates at a speed of 1510 revolutions per minute. The drum provides for precise location and positioning of a data strip. The average rotational delay of the drum is 20 milliseconds.

(2) a load/unload mechanism, for transfer of data strips between the carousel unit and the surface of the strip drum.

(3) a track-selection carriage unit, which supports an optical head incorporating a mirror galvanometer and an objective lens, which can direct a laser light beam onto any selected track region of a data strip.

In the operation of the Laser Recorder Unit (LRU) to load a new data strip, the following take place:

(1) the carousel is positioned under the load/unload station to receive the strip presently mounted on the rotating drum;

(2) the drum is decelerated to stop at a particular point to unload the presently mounted strip into the carousel;

(3) the presently mounted strip is unloaded;

(4) the carousel is rotated to bring the desired strip to the load/unload station; and

(5) the drum is accelerated slowly as it loads the strip and then more rapidly until it reaches 1510 RPM.

These five steps occur in less than 10 seconds, and the procedure could have been directed to either read/record unit. Control now passes to the track-selection carriage unit which moves longitudinally down the drum (across the tracks). When it is positioned approximately over the desired track, the searching of the track is accomplished by the track moving under the optical head at 1510 RPM. Given that it takes 150 milliseconds to access a record once the strip is loaded, and that the drum has an average latency time of 20 milliseconds, then it can be deduced that it takes an average of 130 milliseconds to access a given track. The time to get from track (i) to track (i + 1) is less than 2 milliseconds, while getting from track 1 to track 11,440 requires less than 400 milliseconds. Once a desired record is located, transfer of data then proceeds at 3.4 million bits per second.

The Recorder Control Unit (RCU) consists of a control computer, two word processors, two buffer core memories and associated priority controls, two read/write and error-control subsystems, and necessary I/O control units. The control computer is interfaced directly with the host computer and provides all system control functions. The word-processor programs are loaded and monitored by the control computer and provide the software interface between the data strip, buffer core memories and host computer. Assuming that the host computer commands a record be written in a specific position on a specified file, the control computer loads the word processor with the write program, which includes the absolute track address and the physical record number of the record to be written. The word processor primes the memory address register of the buffer core memory, activates this memory and takes over control of the carriage and galvanometer, directing the laser beam to the proper position for the write operation - a requirement that briefly throws the word processor into the read mode to locate the proper point on the partly filled data strip where the write operation is to begin.

The RCU also establishes and maintains on each strip a directory of all data sets or files, and a directory of all available space. It performs all track allocation for data sets or files and all record address conversions from relative position to absolute track addresses. The error rate for the UNICON is reported to be less than 1 in 10^8 bits.

Figure V-1 is a drawing of the hardware configuration for the UNICON mass memory system that was used in the performance and applications analysis described below. There are two read/write stations each controlled by a separate word processor. The control computer controls the word processors and the flow of information between the staging disks and the host computer. Each read/write station has an I/O channel that is double buffered for effective continuous data flow between the read/write stations and the staging disks.

2. Software Requirements
Precision Instrument Company indicates that it provides the software required to control the data accesses both to and from the data strips as well as providing maintenance programs and system diagnostic routines. Precision Instrument Company also indicates that it supplies the software necessary for any interface requirements.

B. FILE ORGANIZATION

1. Sequential

To store the address of every record that is recorded on the UNICON memory would involve an extremely large VTOC (volume table of contents) to be maintained by a host computer, perhaps as large as 18 X 10^8 bits. Therefore, it is necessary to use the UNICON to store part of the index. For this purpose an STOC (strip table of contents) is recorded in the first track of every strip. Thus, the host computer need only store and address the strip number of a given record. The STOC contains track number and relative address within a track of all records recorded on a strip. The STOC must be loaded into the control computer each time a new strip is mounted.

The method of physically recording data on strips will be considered next. It will be assumed that fixed length records are used and that there is one record per physical block. Several methods are possible, as follows:

The most direct method is to write the data from track 1 sequentially to track 11,440. For archival storage, whether for sequential or random retrieval, this would be the most economic solution. If however, file maintenance is necessary, this method would require all changed and inserted records to be recorded out of sequence on another strip.

Assuming a file maintenance requirement, another method would be to record on track (i) and skip track (i + 1), reserving it for future updates of records in track (i). Expanding on this idea, data would be recorded on the first several tracks of a strip and several tracks would be reserved for updated records, depending upon the activity factor of a data set. If an entire track is not rewritten each time

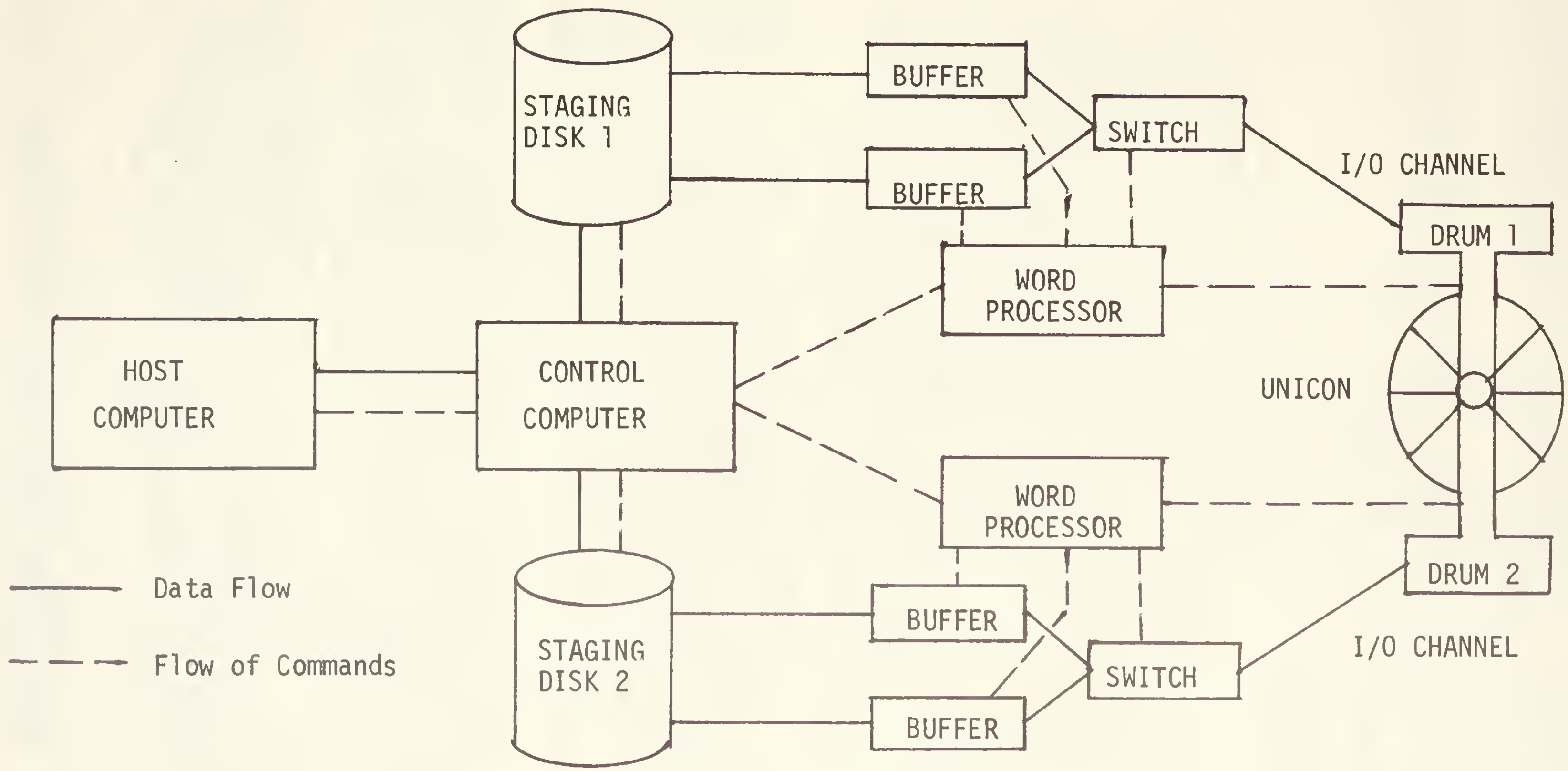


FIGURE V-1 - UNICON HARDWARE CONFIGURATION a record is updated, this method would result in a large amount of track switching. If, however, an entire track is rewritten each time a record is updated, a large amount of blank space would be needed for a file with a high update activity.

Another method would be to record sequentially from track 1 to 11,440, leaving no empty tracks. However, each track would be only partially filled, leaving empty space for updated records. Thus, track switching is greatly reduced.

a. File Maintenance

File maintenance as applied to the UNICON mass memory system and using the third file organization method described above will be considered. The empty space on each track will be referred to as update space. The size of this update space is a function of the file maintenance mode, i.e., batch or demand processing; the number of file changes per day; the life of the strip in days; and the file activity factor. For the present, batch processing and daily file maintenance will be assumed. The life of a strip depends on how much space is allocated for updates and how many days it takes to fill this space. For example, if data is stored for archiving, the strip life would be indefinite. The activity factor refers to the percentage of the records within a file which are accessed or updated. For example, an activity factor of .1 indicates that on the average, every tenth record will be accessed. A further assumption is that both read/write stations will be used.

Each time file maintenance is performed on a strip, a new STOC must be written. Due to the permanent recording of data, this information must be stored in new tracks daily. Thus depending on the life of the strip, some number of tracks must be left vacant at the beginning of each strip. For example, if the life of a strip is 15 days, tracks for storing 15 STOC's must be reserved at the beginning of each strip. When these STOC tracks are filled, the current information must be written on a new strip and the old strip discarded.

When updating a file, records are transferred from the mass memory to the host computer, where records are changed. After reading the STOC, the first step in file updating is to read the entire track containing the desired record or records onto the staging disk (Figure V-1). The STOC, which has been stored in the control computer, has the relative address of all records within the track, so that the desired records can be transferred from the staging disk to the host computer. After the updated records are written on the staging disk, only the updated records are written into the update space of the appropriate track. A change in address is then made for the updated record in the STOC. Finally, when all records on the strip have been updated, the modified STOC is written on the appropriate strip.

When updating records randomly across strips, it is better to read just the desired record rather than the entire track. This is also true for low activity sequential files.

When making insertions, the host computer identifies the strip where records are to be inserted. The strip is mounted and its STOC read into the control computer. The records to be inserted are placed on the staging disk by the host computer. The relative addresses of the records to be inserted are determined by the control computer. The STOC is updated and the inserted record is written into the update space of the appropriate track.

For deletions, the strip containing the file where records are to be deleted has its STOC loaded into the control computer. The relative record addresses of deleted records are removed from the STOC. The STOC is written back onto the appropriate strip.

b. Data Retrieval

It will be assumed that a strip has been used for a number of daily updates and that data is to be retrieved from that strip. The initial step is to load the STOC of the desired strip into the control computer. As a result, access can be made directly to any record on the mounted strip. Assume that records 1 through 6 were initially recorded sequentially and that during file maintenance records 2 and 6 had been updated and moved. As result, the STOC points to positions in the update space for records 2 and 6. To read these six records, the control computer initiates a word processor to control the movement of the optic head. Based on the relative position of a record within a track, the control computer will transfer to the word processor the length of track to be read by the optic head. Thus, record 1 would be read, then the word processor would be given a new location of the track to read record 2. The same procedure occurs to read record 3, except now the control computer sends the length of track for the next 3 sequential records to the word processor. The read procedure would continue in this manner. If a strip contained archival information only, the data would be retrieved sequentially with no need for reading out of sequence.

Another method for data retrieval would be to read the entire track into the buffer and sequentially move one record at a time to the staging disk. This technique would, however, require the buffer to be addressable and at least have the capacity of a track, 2×10^4 bytes. If records are retrieved randomly from the mass memory, the host computer must identify the strip and have the STOC loaded into the control computer. Thus, if records are retrieved randomly across all strips, much time will be lost in loading strips and STOC's.

c. Report Generation

For report generation, a file would be read onto a staging disk and then printed on a line printer or terminal. In order to generate a report of the records in a file, the following takes place. First, the STOC of the desired strip is loaded. Then, the records are read in file order onto the staging disk. A segment of records in a track are read sequentially until reference is made to the update space of the track. If no updates had been made on a track, the entire track, minus the update space, would be read sequentially. Once loaded onto a staging disk, a file is ready to be processed by a line printer.

d. <u>Sorting</u>

The UNICON System is not well suited for sorting because it uses non-erasable storage. It is time consuming to continually move records as required during sorting. Also, the UNICON System has only two read/write stations. This is not adequate for even a two-way merge. Sorting could be accomplished by conventional equipment. By using two staging disks and two magnetic tape drives, the host computer could perform the sorting of data stored in the laser memory. The size of the file that could be sorted would be determined by the number and capacity of staging disks and other peripheral devices.

VI. INTERNATIONAL VIDEO CORPORATION IVC-1000

A. SYSTEM DESCRIPTION

1. Hardware Components

The High Information Density (HID) Recorder or IVC-1000, produced by International Video Corporation is a mass storage device rather than a system. The IVC-1000 includes a video tape recorder transport and basic tape control electronics, but does not include a tape controller (or driver), a data channel or a control computer.

The IVC-1000 uses a helical scan rotating head video recording technique on a one inch wide tape with 7000 feet of tape on a 12-1/2 inch cartridge. Approximately 100 tracks are recorded across the tape at a linear density of over 11,000 bits per inch, resulting in an information density exceeding 1 million bits per square inch. The recorder specification are as follows:

--Tape read/write speed: 6.91 ips (head speed 723 ips)

--Tape search speed: 400 ips bidirectional

--Read/write data transfer rates: 8×10^6 bits/sec

--Tape reel capacity: 9×10^{10} bits (no redundancy)

--Number of recorders for trillion bit capacity on line: 12

--Recording media: standard one inch 3M magnetic video tape

--Capacity per track (physical block size): 120k bits

In addition to the data tracks, there are five longitudinal tracks for status, clock, permanent address, control and a spare. These tracks are recorded at 280 bits/inch. The status track can be read or rewritten at any speed up to 400 ips and could be used as a tally track as is done in the TBM system. The permanent address track is divided into 28 bit segments of which 20 bits in the address of the track and 8 bits are free. Although the address track cannot be rewritten it can be read at any speed up to 400 ips. When the tape is mounted the address and status tracks are normally read and a Volume Table of Contents constructed while the tape is being wound onto the fixed reel and before the first data record is read or written. This feature is optional. The address or status tracks may be searched in either direction at 400 ips. In earlier versions, these longitudinal tracks were spaced across the tape, but in later versions they were placed along the tape edge to decrease the error rate. The error rate in early versions was tested to be 1 bit in 10^7 and is expected to be 1 bit in 10^8 with the edge tracks.

The following timing information is important:

--Time to reach search speed of 400 ips: 1.7 sec

--Time to stop from 400 ips: 2.7 sec

--Time to obtain synchronization and begin reading or writing: 0.6 sec

--Time to rewind tape: 3.5 minutes

--Maximum search time: 215 sec
(1.7 start + .6 synchronization ready for reading + 210 search + 2.7 stop)

--Average search time: 75 sec

--Time for automatic loading: 20 sec

--Time to read (erase) and rewrite a block in place: 5 sec

It is important to note that it takes 0.6 second and one inch of tape to obtain synchronization, during which about 10 tracks are not read.

2. Software

International Video Corporation provides no software. Only a device is provided and the user is expected to supply all support equipment, such as tape controller, data channels, buffers and control computer, as well as all the software for the device.

3. System Firms

Three companies are building systems using IVC recorders, but they did not respond to inquiries.* Transifile has 7 preproduction recorders and is building a system with 13 tracks in each logical block for a capacity of 1.56 million bits per block. They will use a document reader as input and a PDP-15 as control computer. Texas Instruments is building a high reliability system with each track containing only 40k bits but recorded three times, first in true form, then inverted and finally in true form again. By using three recordings and two directional error coding, they hope to obtain an error ratio of 1 bit in 10^{12} to 10^{13}.

System Development Corporation is planning to produce two mass memory storage systems using the IVC recorders. Their systems are now in the conceptual design stage. The basic SDC system, MMSS-1, consists of a pair of IVC-1000 recorders (called MMR-1) interfaced to a 370 selector channel through a 20k word (16 bits) core minicomputer. The minicomputer is connected to the 370 selector channel by a maxi-mini interface and to the recorder by an MMR-1 interface (or controller) as shown in Figure VI-1. The information transfer rates and constraints imposed by each element of the system are as follows:

*The accuracy of information obtained from manufacturers concerning their production plans is not guaranteed.

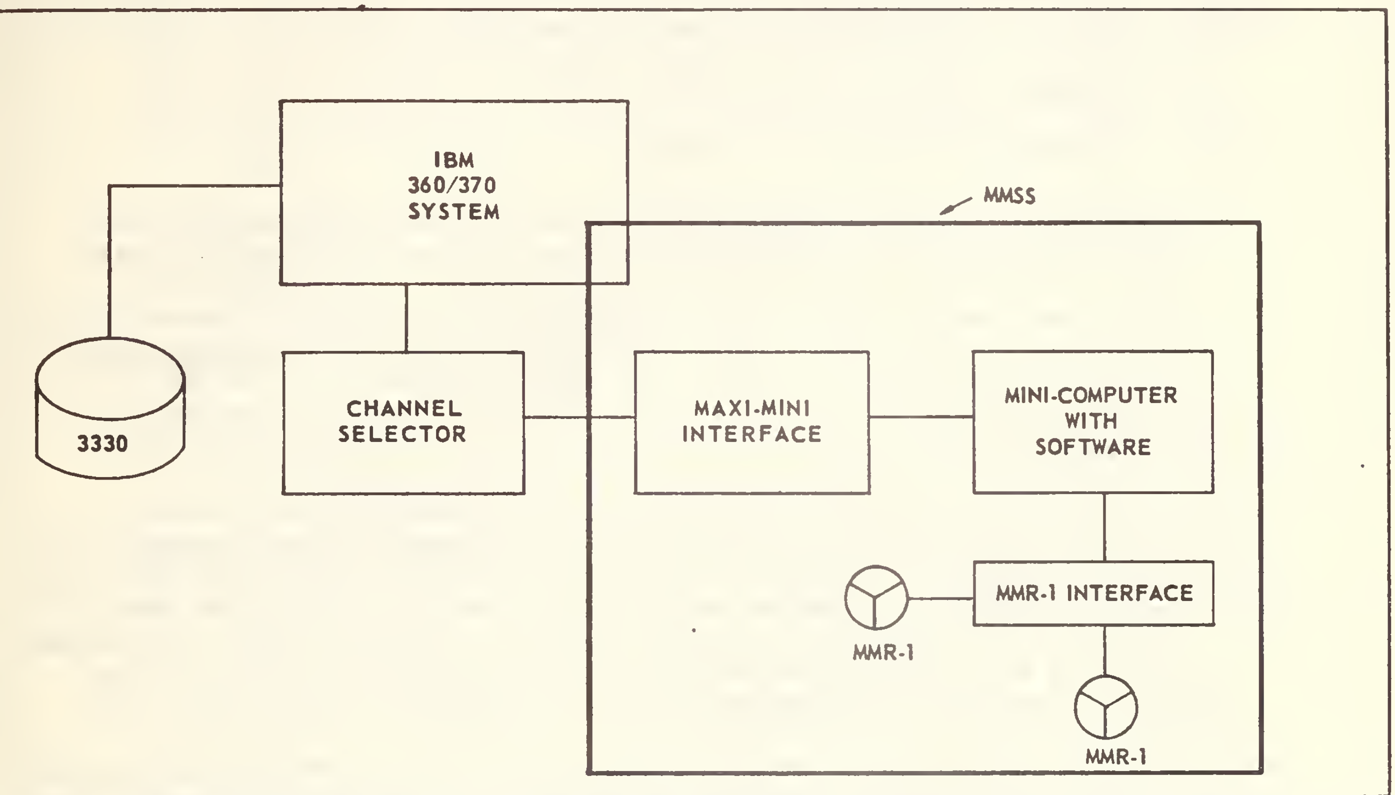


Figure VI-1: SDC's Basic System Configuration (MMSS-1)

a. MMR-1 (IVC-1000): I/O rate of 0.9 MBS (megabytes/second) (burst rate 1.0 MBS),

b. MMR-1 controller: I/O rate of 1.0 MBS,

c. Minicomputer (20k): Effective I.0 rate of 2.0 MBS (1.0 MBS in and 1.0 MBS out),

d. Maxi-mini Interface: I/O rate of 0.45 MBS,

e. Selector Channel: I/O rate of 1.3 MBS,

f. 3330 Disc: I/O rate of 0.8 MBS.

The maxi-mini interface limitation in (d) restricts the transfer rate to half that of the recorder. This can be resolved by using alternate tracks on the recorder for sequential information. This effectively reduces the transfer rate of the recorder to 0.45 megabytes/second. If overlapping and continual reading and writing from the host computer is required, the effective transfer rate must be reduced to 0.22 MBS by using every fourth track. Of course, the other tracks can be used on subsequent passes.

SDC also plans an expanded system, called MMSS-2, that uses two maxi-mini interfaces and two selector channels to connect the minicomputer to the host. Now the limiting transfer rate is 0.8 MBS due to the 3330 disk and this could be accomodated by utilizing 4 out of 5 tracks on the IVC recorders. To obtain 0.9 MBS while reading or 0.45 MBS while reading and writing, a second 3330 disk and its selector channel is used with the MMSS-2. The major components of the SDC system will now be described.

The MMR-1 controller contains all the necessary logic and timing required to interface the MMR-1 and the minicomputer. It permits attachment of at least two and not more than four recorders. The controller controls the search speed of the MMR-1 based on track address comparison. Apparently, only one recorder can be searched at a time. The minicomputer consists of a central processor, core memory for data and instructions, two selector channels which operate at 2 MBS and a Model 33 typewriter. The core memory consists of 16-bit words with a cycle time of 1 microsecond or less. Two modules of 8K words (2 bytes/word) each are required to double buffer the MMR-1 (which transfers 15K byte blocks). One module of 4K words or larger is required for programming, for a total of 20K words. For double buffering of both reading and writing operating, an extra two 8K modules are required for a total of 36K words.

The maxi-mini interface provides the logic and timing necessary to interface the mini computer to the IBM 360/370 channel. The channel can be the 2880 block multiplexor or the 2660 selector channel. The interface appears to the 360/370 channel as a tape controller. The transfer rate of the interface is nominally 0.45 MBS but may be as high as 0.7 MBS.

SDC plans to supply all software, including programs for the mini-computer and modifications to the IBM Operating System (OS). An objective is to make as few changes as possible to OS. Device recognition must be provided in OS, and two new modules will be added to the Queued Sequential Access Method System for moving data between the MMR-1 and the 3330. Since the SDC systems are still in the conceptual design stage, many changes may yet occur in the hardware design. Also, all software must be developed and debugged.

4. Availability and cost

Ten engineering pre-production models of the IVC-1000 have been built; seven have been delivered to Transifile and one to Texas Instruments. The first production models are scheduled for production in November 1972 at the rate of one per week increasing to a rate of five per week. The cost of the complete tape transport is $50,000 to $55,000. Individual tape cartridges cost about $250 each, but with large orders this could be as low as $100. Tape life is 500 read/write operations or 1500 search operations. The cost for video read head is $300, it has an expected life of 10,000 hours, and it requires 30 seconds to replace.

The purchase price for the basic SDS system (MMSS-1) was quoted as $350K to $450K, while the cost of the MMSS-2 was quoted as $450K to $550K.

B. FILE ORGANIZATIONS FOR IVC RECORDERS

1. Introduction

Although sequential, index sequential and random file organizations are considered appropriate for mass storage devices, the sequential file organization is the simplest and most appropriate for IVC recorders. In this method all records are maintained in sequential order such as by the date of entry or alphabetically. Index sequential can be implemented on IVC recorders by using the permanent address and status tracks for index information. (The latter is similar to the Ampex tally track.) A set of indexes can be stored in a table. To locate a record, the index table is first searched for the appropriate index and then that file is searched sequentially for the block or record. Depending on the number of blocks between indexes, the index table may be very large and the time to locate the record may be large. As an alternate to storing indexes in a table, the indexes may be calculated by some randomizing formula. This is sometimes called the calculated index sequential file organization. For large files on mass storage devices, this technique provides faster access than the table method and does not require a large index table. It does have the disadvantage that some file areas may never be used because there may be gaps in the addresses calculated by the randomizing formula. To address all records on an IVC tape, assuming 1000 byte records, would require a 24-bit address and 2.4×10^8 bits of storage for the addresses. A random file organization is not appropriate for IVC recorders because the average time to position the read head in a random search is about 75 seconds after the address has been obtained.

Only sequential and random file organizations will be considered in this report. For IVC recorders, only sequential file organization is considered appropriate for most applications.

The problem of storing variable length records is the same for mass storage devices as with other storage devices. Mass storage devices have fixed length physical block sizes of from 1.2×10^5 to 10^6 bits. The flexibility of variable length records can be obtained in one of three ways: first, a maximum block size can be reserved for every record;

second, an overflow area can be provided for records which no longer fit a physical block; or third, end of logical record markers can be used. The first method is wasteful of space, if the average record size is much less than the maximum. However, this is not as critical with mass storage devices as with smaller storage capacity devices. The second method may result in a disorganized file structure and long search times, particularly when IVC recorders are used. The third method is the most appropriate for IVC recorders because records are maintained in a sequential order and there is no wasted space other than for the end of record markers.

Major data processing functions for a mass storage system using the IVC recorders are:

1. Update
2. Fetch (or retrieve)
3. Purge
4. Store
5. Generate reports
6. Archive (store for years)
7. Report status of units
8. Report percentage and physical distribution of storage in use.

The following assumptions are made in this section:

- Buffers are addressable to the record level;
- Devices are addressable to the block level;
- Buffers will each hold one block;
- Multiple buffers are available;
- Only a single read or write is allowed for the minimum configuration with minimum on-line storage; and
- Simultaneous read and write operations are allowed for the medium configurations which have extra devices, a controller and channels.

2. Sequential File Updating

a. Method 1: Updating in Place

Updating in place is used for erasable storage only and has the disadvan- tage of being slower than the tape-in, tape-out method, since the read, backspace and write operations are all performed serially without overlap. For low activity files, the time to perform extra non-overlapping steps and then skip to the next active record may be less than the time to transfer all the inactive records. For IVC recorders, the file activity factor must be below 0.0003 before up-dating in place is faster than the tape-in, tape-out, method because the time to update in place is about 3.3 seconds and the time to read a record (of 1000 bytes) is about one millisecond.

Updating in place does not provide automatic file back-up capability. It is necessary to occasionally copy files in order to provide a back-up. Also, variable length records must be placed in an overflow area. In addition, deleting records or inserting smaller records causes the tape to have many unused and wasted spaces.

Updating in place would be applicable for IVC recorders under the following conditions:

- Fixed record size;
- Low file activity; and
- Where the file is small and can be conveniently copied during off-hours for backup purposes.

The steps for updating in place with IVC recorders are as follows:

1) Select the first (next) update transaction.
2) Search the index file for the storage location of the record to be updated.
3) Load the necessary tape (requires operator intervention).
4) Position the read head to the beginning of the block and bring the tape and read-head up to speed and synchronization.
5) Read the block into buffer storage.
6) Obtain the address of the desired record in the buffer.
7) Update the record.
8) Repeat 6, 7 for all other records in that block.
9) Backspace to the beginning of the block.
10) Write the updated block from the buffer onto the device.
11) Go to step 1 (steps 3 and 4 are skipped).

To read the next sequential block (or track) with an IVC recorder requires repositioning of the head because the head coasts about 15 tracks while stopping and requires another 15 tracks to start and obtain synchronization for the next read.

b. Method 2: File-in, File-out Updating
Creating a new master file, or file-in, file-out method of updating, is generally superior to updating in place for the following reasons:

1) Files can be read and rewritten concurrently (assuming a 2 channel configuration), so that no time is lost by rewriting the file.

2) This method is much faster for IVC recorders when the buffering is such that continual reading and writing of large blocks can be used.

3) It is the best method for high file activity factors.

4) It can be used for fixed or variable length records.

5) Deletions do not leave wasted space.

6) A backup file is created automatically. (Actually, three tapes should be used so that if the input tape is destroyed while creating a new master there is still another version on the shelf.)

7) The cost for the extra erasable storage medium is small.

The steps for updating records by the file-in, file-out method are:

1-8) Same as 1 to 8 for updating in place if the first record is in the first block. If not, steps 3, 4, 5 and 9 (below) must be performed for the first block in the file.

9) Write the block from the buffer onto the device.

10) Repeat step 5 and 9 until the block containing the first (next) transaction is encountered.

11) Go to step 1.

c. Method 3: Updating Via a Change File
With this method the updated records are stored out of sequence on another file and periodically merged with the master file. This has the disadvantage of having to search more than one file and probably two tapes when retrieving information. The decision to use this method would depend on the total time to update and maintain the file in desired sequence and to retrieve data when required. It would be most useful for files with low maintenance activity factors and low retrieval rates. This method is quite appropriate for IVC recorders, especially if both the master file and the change file are mounted during retrievals.

d. Method 4: Storing Updated Records in Overflow Area
Storing updated records in an overflow area is a technique used primarily for random access devices or non-erasable storage and is generally not suitable for IVC recorders. It is more often used for erasable storage (primarily discs) where variable length records must be

stored or where records are to be inserted in sequential files (discs have a fixed record length format). Sequential retrieval becomes difficult when records are out of sequence, especially on IVC recorders which have limited random access capability.

e. Sequential Retrieval of Records

The steps for individually retrieving records are essentially the same as for updating in place, except that the desired record is read rather than updated. For IVC recorders, reading may be for a single track, large block, or continuous--depending on buffer size and activity factor. The time for retrieving records by blocks is about half that of updating in place.

For activity factors greater than about 0.001, it is more efficient to use continuous reading for sequential record retrieval. In the SDC system, if tracks are skipped so that simultaneous reading and writing loads the limiting resource (channel or interface) to capacity, retrieval times are the same as updating times. The disadvantage of using this file organization method is that the limiting resource is used to only 50 per cent of capacity during retrieval. For example, the maxi-mini interface between the minicomputer and the host computer channel may limit the throughput to 0.45 mbytes/sec for simultaneous reading and writing, forcing the use of every second track for storing sequential records. During retrieval, the throughput is limited to 0.45 mbytes/sec by the IVC recorder (every second track) and the maxi-mini interface is used to half of its 0.9 mbytes/sec capacity.

If retrieval activity is much greater than update activity, it is advisable to record the sequential file so that the limiting resource is used to capacity during retrieval. In the example above, this means recording on every track, which permits reading at 0.9 mbytes/sec. The disadvantage of this method is that the limiting resource cannot handle simultaneous reading and writing for updating purposes.

f. Random Retrieval

The IVC recorder is generally not suitable for random retrieval because it is basically a sequential device. At a tape search speed of 400 ips it requires 70 seconds to make a random search--certainly not competitive with random access devices. Retrieval time can be reduced somewhat by using several tape drives and distributing the file over the drives.

g. Inserting and Deleting Records

It is not practical to use continuous reading and writing with the file-in, file-out method when inserting and deleting records, because for each inserted or deleted block, both the input and output recorders must be stopped and resynchronized with each other. For example, if a block is to be deleted, the output recorder must be stopped waiting for the next block to be written. It takes so long to stop and restart the output recorder that the input recorder must be stopped to wait for the output recorder to start.

h. Sorting

Since sorting involves many file passes during the merge phase, sorting is slow on IVC recorders because of large stop and start times. Sorting is best accomplished by reading the entire file onto disk or drum and sorting it there. If the file is so large that it will not fit on the disk or drum, smaller file segments can be sorted and subsequently merged.

. CONCLUSIONS

Conclusions concerning the applicability, advantages and limitations of high density, mass storage systems follow.

A. APPLICABILITY

Mass storage devices have high potential for those applications which require a large data base (10^{10} - 10^{12} bits) to be stored on-line. In applications involving frequent on-line processing and a large data base which cannot be stored entirely on-line, a significant amount of time is spent in mounting and dismounting tapes or disc packs. The use of mass storage eliminates or reduces this time loss. By using mass storage to store an entire data base on-line, the cost of magnetic tapes and disc packs and the cost and space required for off-line storage can be significantly reduced.

B. COMPARISON WITH CONVENTIONAL STORAGE DEVICES

On an access time basis, mass storage devices are not competitive with direct access units, such as discs and frums, for direct access processing of inquiries and updates. For example, the average access time to a record on the two mounted strips for UNICON is 150 msec.; the corresponding access times for TBM and IVC are 13.3 sec. and 75 sec., respectively. The average access time of the IBM 2314 is 75 msec. Mass storage units would be appropriate for direct access processing where discs and drums have inadequate on-line storage capacity or where non-erasable storage is desired (UNICON).

Mass storage devices are comparable in performance to conventional storage units for sequential processing. For example, the I/O transfer rates per channel are 425,000 bytes/sec.; 750,000 bytes/sec.; and 1,000,000 bytes/sec. for UNICON, TBM AND IVC, respectively, versus 806,000 bytes/sec. for the IBM 3330. Thus, mass storage devices are competitive with magnetic tapes, discs and drums for sequential file processing applications.

C. SUPPLEMENT TO CONVENTIONAL STORAGE DEVICES

Mass storage devices should be considered as supplements to conventional storage devices for large data base applications, and used as part of a hierarchical storage system, rather than as replacements for existing storage units. In many applications, it would be advantageous to use disc or drum units in tandem with mass storage devices. Files would be transferred from mass storage to disc or drum as required. Once a file is stored on a direct access device, data can be retrieved faster for demand inquiries and updates than would be the case if the mass storage device was accessed directly. The use of this scheme requires some regularity in the processing schedule, such that the need for access to certain files can be predicted. Files which will be required for direct access processing would be transferred from mass storage to disc at the beginning of each computer operating cycle or when the need arises. Inquiries or transactions which do not have critical time requirements would be processed against files which are stored only on mass storage devices.

Mass storage systems could also be used to store large programs and files which are rolled into or out of main storage in time sharing and multi-programming environments. Mass storage devices could also be used to queue jobs at various stages of processing and to store output for batch printing.

D. COST

In addition to the cost of mass storage devices, significant costs are incurred for drivers, core buffers, staging discs and the control computer. This equipment is required in order to implement a mass storage system. Thus, the absolute cost of a nominal mass storage system is higher than the cost of a nominal disc storage system. However, the capacity of a nominal mass storage system is about 100 times that of a nominal disc storage system, resulting in a much lower cost per bit in favor of mass storage. For example, the IVC-1000 and TBM provide 10^{11} bits of storage for an approximate price of $500K. A disc system provides about 10^9 bits of storage at a price of approximately $200K. Thus, the cost per bit for the mass storage systems is about $5 per million bits as compared to $200 per million bits for disk storage.

E. FILE ACTIVITY RATIO

For high file activity ratios, sequential file-in, file-out updating and sequential data retrieval, using batched input, are the best methods for mass storage file maintenance and data retrieval, respectively. For low file activity ratios, updating in place and direct data retrieval, using an address look-up or calculation scheme, are the best methods.

An example of this point is that, for IVC recorders,updating 1000 byte records in place would be faster than sequential file maintenance when the file activity ratio is less than about .0003 (update in place time is 3.3 seconds and the data transfer rate is 10^6 bytes/second).

F. LARGE BLOCK SIZE PROBLEM

A disadvantage of mass storage devices is the large block size (1.2×10^5 bits, 1.5×10^5 bits and 10^6 bits for IVC, UNICON and TBM, respectively) which must be accessed and read in order to locate the logical record of interest. With disc and drum units, much smaller records can be accessed directly. However, the ability to access by logical record in the mass storage buffer partially compensates for the large block size problem.

G. NON-ERASABLE MASS STORAGE

Non-erasable mass storage devices, such as UNICON, are most applicable to archival storage systems or where few changes are made to files. Also, non-erasable devices are not well suited for sorting and merging applications, where there is the requirement to erase and re-write files many times during the merge phase of a sort.

In summary, it appears that there is only one major requirement which would justify the use of a mass storage system and that is the need to store a very large data base entirely on-line. Other considerations, such as direct access time, sequential file processing time, sorting and merging efficiency, ease of programming, etc., do not favor mass storage devices over conventional storage equipment.

II. BIBLIOGRAPHY

1. Becker, C. H., "UNICON Computer Mass Memory System," Proceedings--Fall Joint Computer Conference, 1966, pages 711-716.

2. Brooks, Frederick P. "Mass Memory in Computer Systems," IEEE Transactions On Magnetics, Vol. MAG-5, No. 3, September 1969, pages 635-639.

3. Chen, Di and Tufte, Obert N. "Optical Memories--Now and in the Future," Electronics World, October 1970, pages 56-60.

4. "Conceptual Design Study of a Mass Memory Storage System," System Development Corporation, July 1972.

5. Damron, S., et al, "A Random Access Terabit Magnetic Memory," Fall Joint Computer Conference, 1968, pages 1381-1387.

6. "Example of a Small TBM System Interfaced to an IBM 360/65," Ampex Corporporation, pages 135-172.

7. Gentile, Richard B. and Lucas, Joseph R., Jr., "The TABLON Mass Storage Network," Spring Joint Computer Conference, 1971, pages 345-356.

8. High Information Density (HID) Recorder IVC-1000, International Video Corporation, IVC-336, December 1971.

9. Hunt, R. P., Elser T. and Wolf, I. W., "The Future Role of Magnetooptical Memory Systems," Datamation, April 1970, pages 97-101.

10. IVC High Information Density (HID) Recorder, International Video Corporation, IVC-361, January 1972.

11. Knapp, Morris A. and McIntyre, David E., "Bulk Storage Applications in the Illiac IV System," IEEE Transactions on Magnetics, Vol. MAG. 7, No. 4, December 1971, pages 838-842.

12. Kuehler, J. D. and Kerby, H. R., "A Photo-Digital Mass Storage System," Proceedings--Fall Joint Computer Conference, 1966, pages 735-742.

13. Lohman, R. D., Mezrich, R. S. and Stewart, W. C., "Holographic Mass Memory's Promise: Megabits Accessible in Microseconds," Electronics, January 18, 1971, pages 61-66.

14. McFarland, K. and Hashiguchi, M., "Laser Recording Unit for High Density Permanent Digital Data Storage," Fall Joint Computer Conference, 1968, pages 1369-1380.

15. Newberry, Sterling P., "An Electron Optical Technique for Large-Capacity Random-Access Memories," Proceedings--Fall Joint Computer Conference, 1966, pages 717-728.

16. Riley, Wallace B., "How to Miniaturize Mass Memories," Electronics, February 14, 1972, pages 91-92.

17. Risko, Frank D., "New Horizons for Magnetic Bulk Storage Devices," Proceedings--Fall Joint Computer Conference, 1968, pages 1361-1367.

18. "TBM--A Brief Description of the Terabit Memory System," Ampex Corporation, April 7, 1971.

19. "The TBM Mass Memory System," Ampex Corporation, pages 136-572.

20. Weil, J. W., "An Introduction to Massive Stores," Honeywell Computer Journal, (undated), pages 89-92.

INITIAL DISTRIBUTION LIST

	No. Copies
Fleet Material Support Office Code 964 Mechanicsburg, Pennsylvania 17055	100
Defense Documentation Center (DDC) Cameron Station Alexandria, Virginia 22314	12
Library Naval Postgraduate School Monterey, California 93940	2
Dean of Research Administration Naval Postgraduate School Monterey, California 93940	2
Library Department of Operations Research and Administrative Sciences Naval Postgraduate School Monterey, California 93940	1
W. R. Church Computer Center Naval Postgraduate School Monterey, California 93940	1
Office of Naval Research, Code 437 Information Systems Program Department of the Navy Arlington, Virginia 22217	2
Office of Naval Research Branch Office/Boston 495 Summer Street Boston, Massachusetts 02210	1
Office of Naval Research Branch Office/Chicago 536 South Clark Street Chicago, Illinois 60605	1
Office of Naval Research Branch Office/Pasadena 1030 East Green Street Pasadena, California 91101	1
Director, Naval Research Laboratory ATTN: Library, Code 2029 (ONRL) Washington, D. C. 20390	1

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)

Naval Postgraduate School
Monterey, California 93940

2a. REPORT SECURITY CLASSIFICATION

UNCLASSIFIED

2b. GROUP

3. REPORT TITLE

A Survey And Analysis of High Density Mass Storage Devices and Systems

4. DESCRIPTIVE NOTES (Type of report and, inclusive dates)

Technical Report, 1972

5. AUTHOR(S) (First name, middle initial, last name)

Norman F. Schneidewind
Gordon H. Syms
Thomas L. Grainger
Robert J. Carden

6. REPORT DATE

July 1972

7a. TOTAL NO. OF PAGES

48

7b. NO. OF REFS

20

8a. CONTRACT OR GRANT NO.

b. PROJECT NO. PO-2-2099

c.

d.

9a. ORIGINATOR'S REPORT NUMBER(S)

NPS-55SS72071A

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY

Fleet Material Support Office

13. ABSTRACT A survey and analysis has been made of high density mass storage systems for the Navy Fleet Material Support Office. The purpose of the project was to survey mass storage devices and systems and to select several devices for detailed analysis. Representative devices were analyzed in order to determine their suitability for various file management functions. The major conclusions of the study are the following:

1. Mass storage devices have high potential for those applications which have a requirement to store a large data base (10^{10}-10^{12} bits) on-line.
2. Mass storage devices should be considered as supplements to conventional storage devices for large data base applications, and used as part of a hierarchical storage system, rather than as replacements for conventional storage equipment.
3. Mass storage devices are not competitive with conventional storage equipment for direct access processing.
4. Erasable mass storage devices are competitive with conventional storage equipment for sequential file processing.
5. Non-erasable mass storage devices are inappropriate for high activity file processing but can be employed to advantage in archival storage applications.
6. As in the case of conventional storage units, the file activity ratio is a prime consideration in the selection of a file processing technique for mass storage. Low activity ratios favor address look-up or calculation and direct file access. High file activity ratios favor batched input and sequential file access.

KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
High Density						
Recording						
Laser						
Magnetic						
Optical						
Holographic						
Electron Beam						
Video Tape						
Mass Storage						
Storage						
Peripheral						